Translating Principles Into Practice in Regulation Theory

Jean-Jacques Laffont

Working Paper N° 1

(March 1999)

CEER
Centro de Estudios Económicos de la Regulación
Instituto de Economía, Universidad Argentina de la Empresa
Chile 1142, 1° piso
(1098) Buenos Aires, Argentina
Telephone: 54-11-43797693
Fax: 54-11-43797588

(Please see the back pages of this paper for a listing of current working paper titles and subscription ordering information).

The Centro de Estudios de Economía de la Regulación (CEER), is an organization engaged in the analysis of the public regulation of utilities. The CEER is supported by the World Bank, the Regulatory Agencies of the Public Utilities of Telecomunicacions, Power and Gas of Argentina, and the Universidad Argentina de la Empresa (Buenos Aires), where the CEER works.

CEER Working Paper Series

Translating Principles Into Practice in Regulation Theory

Jean-Jacques Laffont

Working Paper N° 1
(March 1999)
JEL N°: D4, L8, L5

Abstract: Regulators must understand basic principles of regulation, to explain with credibility their decissions, to asess the universality of the proposals, to develop new solutions, etc. For the relevance of economic theory, economists must fight to include in their frameworks enough constraints, as regulators face in practice, including technology constraints, even informational constraints, bureaucratic and political ones. Debates on regulation are turning around the world more important, defining the discussion frontier about the role of government in the economy.

Affiliation of the author:
Jean-Jacques Laffont, Université de Toulouse.

CEER
Centro de Estudios Económicos de la Regulación
Instituto de Economía, Universidad Argentina de la Empresa
Chile 1142, 1° piso
(1098) Buenos Aires, Argentina
Telephone: 54-11-43797693
Fax: 54-11-43797588

Translating Principles Into Practice in Regulation Theory

Jean-Jacques Laffont

Conference Held at Universidad Argentina de la Empresa, Buenos Aires, November 11, 1997

It is a great pleasure for me to visit the Economic Regulation Research Center, and it is a great honor to be able to share with you some thoughts about the theory of regulation and its translation into practice.

My discussions with the European Commission, various governmental officials and the World Bank staff have increased my belief of the need for regulators to understand economic principles of regulation to be able to credibly explain their decisions, determine whether one regulator's solution can be applied elsewhere, develop new solutions, etc.

It has also convinced me that, for economic theory to be relevant, it must strive for including in its framework as many constraints as regulators are facing, not only technological, but also informational, admininistrative and political ones.

The debates about regulation are becoming all over the world the frontier of the fundamental questioning of the role of government in the economy. It has replaced the old confrontation between communism and capitalism. It is essential that our discussions be guided by all the available knowledge. The Center for Regulatory Economics Studies of UADE is already playing a highly socially valuable role in bringing together regulators and academics, and I am happy to participate in this process.

Given that my work is essentially theoretical, I conceived my talk as a dialog between theory and the practical tasks facing those (regulators, firms, officials, academics) involved in regulation.

I will proceed as follows.

First, I will stress the main theoretical idea that economics has used to renew its vision of regulation, namely the role of asymmetric information in the allocation of resources.

Then, I will take a few examples of debates which have been fruitfully reconsidered in this new light:

1. What is a good regulation of a monopolized activity?

2. What should be liberalized?

3. How to deal with the interface of regulation and competition?

Economists became aware of the key role played by asymmetric information in the allocation of resources at the beginning of the seventies.

Several contributors, including the recent Nobel laureates, James Mirrlees and William Vickrey, showed that the strategic behavior of economic agents endowed with private information was putting more constraints on the allocation of resources than economists used to consider, namely the technological and accounting constraints.

It was not enough for an allocation of resources to be technologically possible to be actually feasible. Incentive constraints, which reflect the strategic behavior of agents endowed with private information, had to be satisfied for an allocation to be really feasible, say, incentive feasible. The precise modeling of these constraints is a great achievement of modern economics.

Simultaneously, the study of competitive markets --insurance markets, labor markets, second hand markets-- revealed strange behavior for economists trained in classical economics. Competitive equilibria were inefficient (sometimes even did not exist).

Both our normative and positive concepts clearly had to be renovated.

This *prise de conscience* prompted economists to reconsider the basic notions of exchange and contractual relationship. The theory of contracts, sometimes called the principal-agent theory, became the cornerstone of economic research.

The main initial theoretical ideas were developed in the context of the free rider problem for public goods, the theory of taxation and the theory of monopoly pricing.

It is only in 1979 that Loeb and Magat proposed to view regulation as a principal-agent relationship in which a regulator -the principal- attempts to control a firm -the agent- assumed to be a natural monopoly They emphasized that the main difficulty was the regulator's lack of information about the regulated firm.

Their main achievement was to show that, if no attention was paid to redistributive questions or to the cost of public funds, optimal full information regulation was achievable by the use of incentive mechanisms imported from the resolution of the free rider problem: the so-called Clarke-Groves-Vickrey mechanisms.

Intuitively, the regulator should simply give a transfer equal (up to a constant) to the social value of the regulated firm's production. The firm would then maximize the difference between the social value and the cost of production, achieving, in an incentive compatible way, the first best allocation of resources. Exit the problem of regulation.

Why then were so many people involved in complex regulations and many complaining about the inefficiencies of the widespread rate of return regulation or the inefficiencies attributed to public firms?

In 1982, Baron-Myerson showed that there was a trade-off between efficiency and the unavoidable informational rents that must be given up to the regulated firm, when the regulator wants a project to be realized but does not know the cost of the regulated firm. The intuition for this rent is that a more efficient firm can mimic an inefficient one, i.e. achieve the same production level at a lower cost.

If those rents are costly for society because of distributional reasons -for example if the firms' profits are less valued socially than consumers' welfare-, optimal regulation requires to deviate from efficiency to mitigate the informational rents. As the rents are proportional to the level of production, the only way to capture them is to inflate the price above Ramsey pricing. Baron-Myerson's model rules out regulatory monitoring of accounting data and therefore rules out from the analysis earnings sharing schemes, rate of return regulation, price cap with profitability based ratchet, etc.

In reaction, in 1986, J. Tirole and myself introduced cost or profit observability, in a model with a richer asymmetry of information -both on the technology of the firm and on the unobservable cost reducing efforts of the firm.

When observing, say, high cost the regulator does not know if it is because the firm is an inefficient one or if it has exerted little effort. This model turned out to be a reasonable caricature of the regulation problem faced in industries like the telecommunications and electricity industries; and, a key for theoretical work, turned out to be easily manageable.

How does optimal regulation look like in such a world? Roughly, as the offer of a menu of cost reimbursement rules in which firms would self-select themselves -the most efficient firms choosing high powered schemes close to fixed payments and less efficient firms choosing partial reimbursement of their costs, possibly full reimbursement of costs, offer combined with Ramsey pricing of the goods calculated with the technological marginal costs induced by the selected cost reimbursement rules. An important difference with Baron-Myerson is that we have now two instruments, cost reimbursement rules and pricing, so that there is not necessarily a need for distorting prices to decrease informational rents.

What was then the revolution of applied regulation? Criticisms of the lack of cost minimization embedded in rate of return regulation led in England and in the USA to the price cap proposal. Basically a move from cost plus regulation to fixed price regulation.

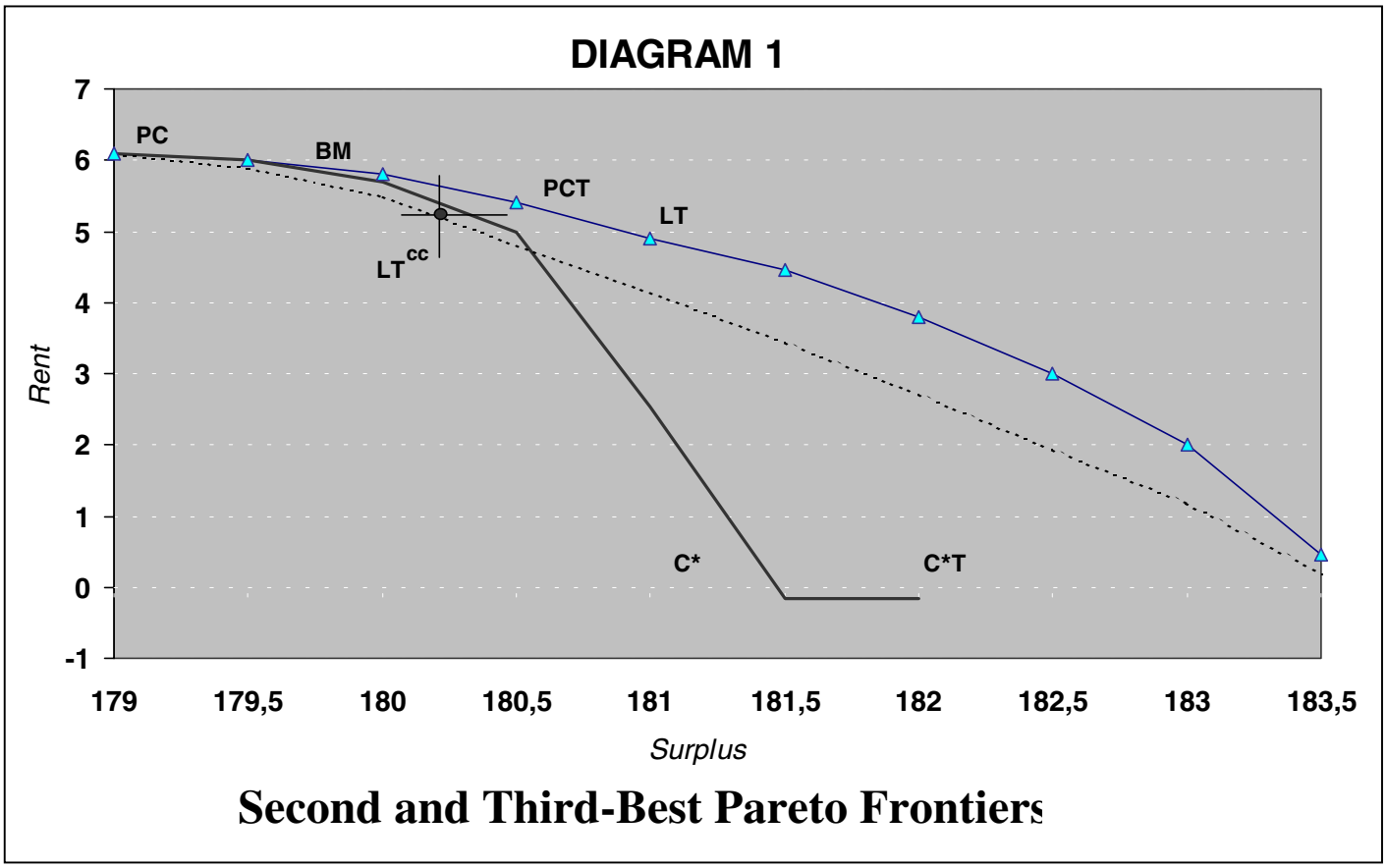
What our theory was saying was that there is no “free lunch” and that the efficiency gains promised by price cap regulation had to be balanced with the unavoidable higher informational rents that would have to be given up.

This major insight from theory was eventually recognized by the US Federal Communications Commission who proposed to redistribute part of the profits of the firms when they were considered to be excessive. This weakened the incentives for cost minimization but reduced the informational rent.

Simulations show that such a combination, appropriately optimized, yields results very close to the theoretical optimal regulation under asymmetric information.

Of course, the dynamic world in which we live makes the story more complex, but this is an example where the logic of economic theory preceded practical regulation.

We then devoted a lot of effort to investigate the implications for regulation of the lack of commitment ability of regulators.



What theory suggests is that optimal regulation should commit ex ante not to use the information about the firm acquired over time. In other words, price caps should be flexible, i.e. indexed but infinitely lived. This allows the regulator to spread evenly over time the distortions called for by the costly informational rents.

In practice, this is not possible of course. Either, the cap is too high and the firm will eventually make enormous profits that will be opposed by consumers; or, it is too low and the firm goes bankrupt. In both cases, renegotiation, which takes into account information about costs revealed during the relationship, will occur.

Anticipating this renegotiation and a potential expropriation of its rents from then on, the firm will behave in such a way that the regulator cannot discover completely its private information to preserve its future informational rents.

This is the famous ratchet effect, which is a fundamental implication of the impossibility in most circumstances to commit not to renegotiate. More research is needed to study the ratchet

effect in widely used mechanisms such as price cap with or without partial redistribution of profits.

A complement to the improvement of cost-reimbursement rules is the introduction, whenever it is possible, of competition. The rapid technological progress occurring in some regulated industries, in particular in telecommunications, led to a questioning of the natural monopoly foundations of these activities, monopolized all over the world. The idea was then to decompose those activities into segments to isolate those incurring high fixed costs in order to open the others, such as value added services or long distance telecommunications, to competition. Similarly, generation of electricity or gas has for some time been considered to be a potentially competitive industry.

Let me stress here the contribution of the economics of information to this debate. Under asymmetric information, one must think in terms of generalized costs, which include informational rents. Then, even if there exist costs of duplication, incurring those costs may be socially valuable because of the yardstick competition effect it entails, decreasing informational rents faster than it increases fixed costs.

To defend monopoly provision, it is now necessary to argue not only that there exist fixed costs, but that these fixed costs are large relative to the potential gains of yardstick competition.

In a simulation exercise based on the Gabel and Kennet local exchange cost optimization model (LECOM), I, together with F. Gasmi and W. Sharkey, have shown that, for a network of 120.000 subscribers in an area of 150 Km², and an asymmetry of information of the order of 20 to 30% of costs, the duplication of fixed costs is not valuable even if the yardstick competition effect is maximal, that is, destroys informational rents in a duopoly.

The liberalization of the segments which are potentially competitive in this new sense leads to two types of issues that we will consider sequentially.

First, to the extent that there still exist bottlenecks, local exchange in telecommunications, transport facilities --pipelines-- electric grid, tracks and stations, access charges for using these bottlenecks must be determined.

Second, despite the appearance of several competitors in the non monopolized segments, competition remains often limited and raises the issue of which form competitive policy must take in these liberalized network industries.

The setting of access charges is a highly difficult exercise. The stakes are high -40 to 50% of the Regional Bell Operating Companies revenues consist in access charges-; the ability of the incumbent to obtain reasonable rates of return on its investment in infrastructures and therefore to develop further the network rests on these charges; the ability of entrants to successfully enter when they are more efficient than the incumbent as well as the temptation for other entrants to inefficiently bypass existing networks also depends on these charges.

The debate among economists on what should be those access charges is highly relevant for applied regulation. For a while the Baumol-Willig efficient component pricing rule (ECPR) dominated the debate. The focus here was on efficient entry by requiring the access price to

equate the opportunity cost of the incumbent. With Jean Tirole, on the basis of the normative theory of optimal pricing under incomplete information, which yields some form of Ramsey pricing, we suggested that ECPR was based on too simple a view of entry --the contestability paradigm-- and was overlooking other dimensions of the problem: information, fixed costs, heterogeneity of firms. Standard perfectly contestable market theory assumes that all firms (incumbents and entrants) face identical cost functions and the same demand function in the competitive segment. Because it also abstracts from incentive issues, it provides no reason for having more than a single firm in the competitive segment in the absence of diminishing returns in that segment. So in the standard framework of cost symmetry, there is no clear motivation for entry. Proponents of ECPR therefore must have in mind an extension of perfectly contestable market theory in which entrants are (at least sometimes) more efficient than the telephone operator (to take the example of telecommunications) in the competitive segment. On the other hand, ECPR implies that the telephone operator (in equilibrium) supplies only access and exits the competitive segment. This does not mean that the access price is completely irrelevant, because it defines who (TO or competitors) will corner the potentially competitive segment. Rather, the prediction of the rule is too stark.

The optimal access price can be derived from optimal Ramsey pricing. It entails a margin over marginal cost which depends on the size of fixed costs in the industry and on the superelasticity of the competitive segment asking for access.

Note that this rule requires discrimination across markets with different elasticities and that ECPR, which coincides with the optimal access rule only in some exceptional cases, also has this feature of being partly demand based.

This demand based approach of access pricing recommended by theory is often opposed by practitioneers. We have two reactions to this conflict.

It is true that there might be some political economic foundations to the no discrimination requirement. Indeed, political interference to favor a particular segment benefiting from a particularly high elasticity is then more difficult.

This argument should not be confused with the less serious one according to which elasticities are just hard to compute.

This leads me to our proposal to implement usage-based access prices in the global price cap policy.

Such a policy includes, in the price cap, both access revenues and other revenues. That is, the intermediate good --access-- is treated as a final good. Weights used in the computation of the price cap are exogenously determined and should be proportional to the forecasted quantities of the associated goods.

An appropriate price cap induces a firm to select the optimal Ramsey structure as long as all goods --including here access goods-- are included in the definition of the cap and the weights are exogenously fixed at the level of output that will be realized. Of course, the exogeneity of weights is a qualifier to this encouraging result as weights based on realizations of output create some difficulties. By not including the access charge in the price cap, a standard price cap de

facto subsidizes non-competitive segments to the detriment of competitive ones. To see this, suppose one adopts instead a partial price cap on retail prices, together with ECPR on access. A TO which maximizes profit will bias its rate structure relative to the Ramsey optimum. Long distance prices and the access charge are too high while local calls are too cheap.

We also argue that global price caps substantially reduce the firm's incentive to foreclose its rivals through non-price methods (such as delays in interconnection, refusal to unbundle, or costly technical requirements). The possibility of foreclusion has been hotly debated in the context of the Regional Bell Operating Companies' entry into long distance in the period preceding the 1996 Telecommunications Act, and figures prominently in several disputes in other countries. Intuitively, under a global price cap, the regulated firm can elect to earn most of its income on interconnection and it then has no incentive to reduce its demand for access. In contrast, current regulations tend to unevenly put more pressure on access charges than on prices of services in competitive segments, and encourage foreclusion. Second, global price caps eliminate another asymmetry in the treatment of product lines that is frequently encountered in existing fragmented regulations. Namely, current regulations provide different incentives for cost reduction or profit enhancement in different product lines. By not balancing incentives properly, they encourage cross subsidies.

Note that I give the phrase "price cap" a more general meaning than contemporary regulatory usage. A price cap is logically consistent with profit sharing rules, although its usage has been restricted to situations in which the regulated firm is (theoretically) residual claimant for its profit. That is, once a price cap has been set, any profit sharing mechanism can be superimposed without affecting the implementation of the structure of Ramsey prices.

To sum up, the global price cap can avoid inefficiencies created by differential incentives in final product revenues and access pricing revenues, would lead approximately to the optimal Ramsey prices and would minimize political interference. The worries about predatory price squeezes could be addressed with price bounds such as the ones derived from ECPR.

The currently dominant paradigm for setting access charges for elements of telephone networks (such as local networks, transmission facilities, switches) is marginal cost, or more precisely "Forward-Looking Long-Run Incremental Cost". To compute this cost, the regulator will perform a cost optimization based on an engineering model, compute an economic depreciation (physical depreciation plus technological and economic progress), and forecast the likely future usage of the elements. This paradigm has hardly been used to date, but was imposed by Congress in the 1996 US Telecommunications Act, adopted in 1995 by the Office of Telecommunications in the UK for 1997 on, and dominates the thinking in Continental Europe.

This dominant paradigm raises several concerns.

First, it does not obey basic economic principles. While marginal cost pricing is the proper benchmark in a competitive industry, the very concept of a bottleneck and the need for regulation stem from the existence of a large fixed cost. Networks involving large fixed costs would never be built if their owners were allowed to charge only marginal cost. This consideration must for instance be born in mind at a time when experts and politicians envision the building of a very costly information superhighway.

Second, and relatedly, marginal cost pricing of access prevents incumbents from making money on the access business and gives them incentives to report their untapped market power on that segment to the competitive segments, by denying access to their rivals through nonprice methods. This possibility calls for close regulatory monitoring and interference with the operators' business judgment.

Third, the computation of marginal costs leaves regulators in charge of setting individual prices and is discretionary. The last two points raise the concern that, beyond the liberalization and free market rhetorics, one may be creating an environment that will lead to heavy-handed regulation.

Besides these one-way access pricing issues, a new set of interconnection issues will develop if competition of integrated networks develops or the deregulation of telecommunications succeeds in creating local competition. The 1996 Telecommunications Act aims at promoting local competition in order to break the local monopoly position of the Regional Bell Operating Companies. It envisions three types of local entry:

- facilities based entry by mobile operators or by fixed-link operators,
- resale entry, through the resale of the local loop by the Baby Bells,
- mixed entry, through unbundling whereby entrants lease some facilities and provide others (e.g., switches).

Regardless of the exact mode of entry, the various networks will need to interconnect to have access to their mutual bottlenecks --the final access to the consumer. Many experts argue that regulation should end once local competition has developed and that regulation should be replaced by standard competition policy (as is the case in New Zealand, where regulatory agencies have been abolished). This (perhaps correct) view is unfortunately not supported by any economic analysis. A tempting analogy with other industries may suggest that competition will yield a socially desirable outcome. This analogy however ignores the fact that interconnection requires an agreement among competitors. This feature, which also arises in various guises in the connection of fixed link and mobile operators, in credit cards and ATM (automatic teller mechanisms) markets and in cross-licensing arrangements raises two concerns. The first is that strong players might refuse to enter interconnection agreements with smaller ones. The second is that the strong players could use the interconnection agreements among themselves as an instrument of tacit collusion in the retail markets.

Consider a mature industry where two telecommunication operators provide a full range of services, and assume that the two operators have symmetric cost structures, have full coverage and are interconnected. The two networks will compete for subscribers but have to agree on mutual access conditions for cross-networks communications. If access charges are given and symmetric, and if networks compete in linear prices, each operator will take into account the impact of his retail prices on its market share, on the revenue generated by its subscribers and on the access revenue or deficit generated by cross-networks communications. The new effect here is that an increase in the retail price increases the access revenue, because it decreases its off-net calls.

This access revenue effect creates a factor of instability of competition between the two networks. If the access charge is high, this effect pushes the final price upwards, possibly higher than the monopoly level, in any shared market equilibrium. But when the prices are very high, and if networks are sufficiently close substitutes, each operator has an incentive to undercut its rival and corner the market. But clearly a cornered market configuration cannot be an equilibrium either.

But this access revenue effect makes also the access charge an instrument of tacit collusion. As we just noted, because of this effect, at a share market equilibrium, the retail equilibrium price increases with the access charge. There is a level of access charge which induces the monopoly price as a non-cooperative equilibrium.

Allowing two part tariffs mitigates this collusive effect. Allowing price discrimination based on call termination creates a wedge between on-net and off-net prices which is detrimental to consumption efficiency, but may intensify competition.

Imposing reciprocity in the setting of access charges may be a particularly good idea when entry and coverage are at stake. However, reciprocity will not in general suffice to prevent entry deterrence. If the incumbent can discriminate between its monopoly territory and the competitive territory, it can still effectively block entry even under a reciprocity rule by insisting on a high access charge, and a high retail price for off-net communications. By discouraging its own subscribers to call the other network -a strategy the entrant cannot follow if its coverage is small-, it avoids paying high access charges to its rival and blocks entry even though formally the access charges are reciprocal.

I hope these examples clearly show that regulation, or at least a sophisticated competitive policy, will stay with us even in the world of competing networks. It is now time to conclude.

It must be acknowledged, despite some of my modest claims, that theory trails practice in most cases, as usual. To provide even with some delay a reasonably rigorous framework to organize our thoughts on regulation remains, I hope, a useful endeavor.

It helps better understand the lessons to be drawn from the frontier experiments, to make other countries, in particular the developing countries, benefit fully from these experiments.

It helps raise some criticisms against proposals too quickly accepted by a community in search for solutions.

CEER Working Paper Series

Number	Author(s)	Title	Date (mm/yy)
1	Laffont, Jean Jacques	Translating Principle Into Practice	03/1999